

全变差图像恢复的自适应步长梯度投影算法

张本鑫^{1,2} 朱志斌^{3,4}

摘要 针对图像去噪问题, 本文基于全变差对偶公式提出一个新的梯度投影算法. 算法采用改进的非单调线搜索和自适应 BB (Barzilai-Borwein) 步长, 有效地改善了 Chambolle 梯度投影算法收敛慢的缺点. 数值结果表明新算法优于一些已有的梯度投影算法.

关键词 梯度投影, 全变差, 自适应步长, 改进的线搜索, 图像恢复

引用格式 张本鑫, 朱志斌. 全变差图像恢复的自适应步长梯度投影算法. 自动化学报, 2016, 42(9): 1347–1355

DOI 10.16383/j.aas.2016.c150146

Gradient Projection Algorithm for Total Variation Image Restoration by Adaptive Steplength Selection Rules

ZHANG Ben-Xin^{1,2} ZHU Zhi-Bin^{3,4}

Abstract We propose a new gradient projection algorithm for image denoising based on the dual of total variation. The new method exploits nonmonotone line-search and adaptive steplength selection based on strategies for alternation of the well-known Barzilai-Borwein rules. The proposed method is much faster than the Chambolle's gradient projection algorithm. Numerical results illustrate the efficiency of this method.

Key words Gradient projection, total variation, adaptive steplength selection, new line search, image restoration

Citation Zhang Ben-Xin, Zhu Zhi-Bin. Gradient projection algorithm for total variation image restoration by adaptive steplength selection rules. *Acta Automatica Sinica*, 2016, 42(9): 1347–1355

自从 Rudin, Osher 和 Fatemi (ROF) 第一次提出全变差 (Total variation, TV) 去噪模型^[1], TV

模型已经成为图像处理领域中非常成功的技术, 在图像恢复、去模糊、重建、修复等^[2–5] 方面得到广泛应用. ROF 模型可以保留图像的不连续边界同时用下面的最小化函数去除噪音:

$$\min_u P(u) := \int_{\Omega} |\nabla u(x)| dx + \frac{\lambda}{2} \|u - f\|_2^2 \quad (1)$$

其中 ∇ 是梯度算子, $\int_{\Omega} |\nabla u(x)| dx$ 表示 u 的全变差. Ω 是图像的定义域, 一般情况下 Ω 是 \mathbf{R}^2 上的矩形. $f: \Omega \rightarrow \mathbf{R}$ 是降质的图像. 通过求解问题 (1) 得到最优点 u 是恢复的图像. λ 为正 regularization 参数, 用于调节数据拟合项与正则项之间的权重. $\|\cdot\|$ 表示 \mathbf{R}^2 上的欧式范数. u 的全变差有如下的对偶形式:

$$\int_{\Omega} |\nabla u| dx = \max_{|\omega| \leq 1} \int_{\Omega} \nabla u \cdot \omega = \max_{|\omega| \leq 1} \int_{\Omega} -u \nabla \cdot \omega \quad (2)$$

其中 $\omega: \Omega \rightarrow \mathbf{R}^2$ 是对偶变量. $\nabla \cdot$ 是散度算子. 最近许多学者提出一些关于对偶 ROF 公式的算法. 由于在对偶的情况下, 不需要进行光滑化处理. 因此可以直接得到原问题的最优解. Chan 等提出对偶的想法^[6], 并用牛顿法求解对偶形式的 ROF 模型. 因此他们的方法具有局部二次收敛速度. 但算法需要计算矩阵的逆. Chambolle 提出了梯度投影下降算

收稿日期 2015-03-24 录用日期 2016-04-28
Manuscript received March 24, 2015; accepted April 28, 2016
国家自然科学基金 (11361018, 11461015), 广西省自然科学基金 (2014GXNSFFA118001), 广西自动检测技术与仪器重点实验室基金 (YQ15112, YQ16112), 广西高校科研一般项目 (KY2016YB167), 桂林市科技攻关项目 (20140127-2), 广西和桂林电子科技大学研究生教育创新计划项目 (YJ CXB201502) 资助
Supported by National Natural Science Foundation of China (11361018, 11461015), Natural Science Foundation of Guangxi Province (2014GXNSFFA118001), Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ15112, YQ16112), Guangxi Education Scientific Research Program (KY2016YB167), Guilin Science and Technology Project (20140127-2), Innovation Project of Guangxi and Guilin University of Electronic Technology Graduate Education (YJ CXB201502)

本文责任编辑 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 桂林电子科技大学电子工程与自动化学院 桂林 541004 2. 桂林电子科技大学广西自动检测技术与仪器重点实验室 桂林 541004 3. 桂林电子科技大学数学与计算科学学院 桂林 541004 4. 桂林电子科技大学广西高校数据分析与计算重点实验室 桂林 541004
1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004 2. Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004 3. School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin 541004 4. Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation, Guilin University of Electronic Technology, Guilin 541004

法^[7-8], 它能快速收敛于中等精度而被广泛应用. 随后许多学者提出了基于 Chambolle 全变差最小化算法. 基于二阶广义全变差正则项, 文献 [9] 提出了模糊图像恢复的分裂 Bregman 算法. 二阶全变差可能需要较多的计算这会导致算法的 CPU 时间增多. 文献 [10] 提出了一个结合 BB (Barzilai-Borwein) 步长的谱共轭梯度投影算法求解图像去噪问题, 该算法使用了单调线搜索, 这可能会限制 BB 步长的数值表现. Xiao 等提出了一个求解 l_1 非光滑问题的非单调 BB 梯度算法^[11], 数值结果表明可以和比较著名的算法相媲美. 文献 [12-13] 将交替使用 BB 步长的投影梯度算法应用到压缩感知和非负矩阵的分解上, 并取得了较好的效果. 文献 [14] 在子算法中结合 BB 步长, 有效地求解了核范数极小化问题. 文献 [15] 提出了求解全变差图像恢复问题的非单调梯度投影算法. 将 Chambolle 的最小化全变差方法拓展到非单调下降方法, 先用 BB 步长代替 Chambolle 投影算法的常数步长, 再用 Dai 等提出的自适应非单调线搜索^[16] 保证算法的全局收敛性.

在高精度要求时, 数值试验表明, 交替使用 BB 步长比单独使用某一个 BB 公式的效果更好. 因此, 本文提出了一个基于全变差图像恢复自适应步长选择的 BB 梯度投影算法. 为了保证算法的收敛性, 给出一个修正的非单调线搜索. 这个线搜索可以得到一个较大的步长以及较少的计算函数值次数, 因而减少算法迭代的时间.

1 自适应 BB 步长投影算法

Chambolle 在文献 [7] 中得到问题 (1) 的解 u 表示如下:

$$u = f + \pi_{\frac{1}{\lambda}\mathcal{K}}(-f) \tag{3}$$

其中 $\pi_{\frac{1}{\lambda}\mathcal{K}}(\cdot)$ 是在凸集 $\frac{1}{\lambda}\mathcal{K}$ 上的正交投影,

$$\mathcal{K} := \{\nabla \cdot \omega : \omega \in Y, |\omega_{i,j}| \leq 1, \forall i, j = 1, \dots, N\}$$

上式中散度算子 $\nabla \cdot : X \rightarrow Y$ 定义为 $\nabla \cdot = -\nabla^*$, 即对任意的 $\omega \in Y$ 和 $u \in X$, $\langle -\nabla \cdot \omega, u \rangle = \langle \omega, \nabla u \rangle$. 因此求取式 (1) 的解取决于计算非线性映射 $\pi_{\frac{1}{\lambda}\mathcal{K}}(-f)$, 即等价于求解下面的约束优化问题:

$$\begin{aligned} & \min_{\omega \in \mathcal{K}} \|f + \frac{1}{\lambda} \nabla \cdot \omega\|_X^2 \\ & \text{s.t. } \mathcal{K} := \{\omega : \omega \in Y, |\omega_{i,j}| \leq 1, \forall i, j = 1, \dots, N\} \end{aligned} \tag{4}$$

对偶问题 (4) 的 KKT (Karush-Kuhn-Tucker) 条件为:

$$-(\nabla(\nabla \cdot \omega + \lambda f))_{i,j} + \alpha_{i,j} \omega_{i,j} = 0$$

其中当 $|\omega_{i,j}| = 1$ 时, $\alpha_{i,j} > 0$. 当 $|\omega_{i,j}| < 1$ 时, $\alpha_{i,j} = 0$. 在后一种情况下, $(\nabla(\nabla \cdot \omega + \lambda f))_{i,j} = 0$. 故, 综合可知拉格朗日乘子 $\alpha \in X$ 满足

$$\alpha_{i,j} = |(\nabla(\nabla \cdot \omega + \lambda f))_{i,j}|, \quad \forall i, j = 1, \dots, N$$

由上述分析可知问题 (4) 的欧拉-拉格朗日公式为

$$-(\nabla(\nabla \cdot \omega + \lambda f))_{i,j} + |(\nabla(\nabla \cdot \omega + \lambda f))_{i,j}| \omega_{i,j} = 0 \tag{5}$$

因此, Chambolle 提出了一个半隐式梯度下降算法:

$$\omega^{k+1} = \omega^k - \tau(-\nabla(\nabla \cdot \omega^k + \lambda f) + |\nabla(\nabla \cdot \omega^k + \lambda f)| \omega^{k+1}) \tag{6}$$

或

$$\omega_{i,j}^{k+1} = \frac{\omega_{i,j}^k + \tau(\nabla(\nabla \cdot \omega^k + \lambda f))_{i,j}}{1 + \tau|(\nabla(\nabla \cdot \omega^k + \lambda f))_{i,j}|} \tag{7}$$

其中 $\tau > 0$ 是一个时间步长. 显然, 在 Chambolle 的投影算法中只要初始条件满足 $|\omega| \leq 1$, 那么在迭代过程中 ω 自然满足约束条件. 为了保证算法 (7) 的收敛性, Chambolle 给出了下面的充分条件.

定理 1. 假设参数 $0 < \tau \leq 1/8$. 那么当 $k \rightarrow \infty$ 时, $\frac{1}{\lambda} \nabla \cdot \omega^k$ 收敛于 $\pi_{\frac{1}{\lambda}\mathcal{K}}(-f)$.

在文献 [8] 中, Chambolle 建议用一个简单的投影梯度下降方法代替式 (7), 即

$$\omega_{i,j}^{k+1} = \frac{\omega_{i,j}^k + \tau(\nabla(\nabla \cdot \omega^k + \lambda f))_{i,j}}{\max\{1, |\omega_{i,j}^k + \tau(\nabla(\nabla \cdot \omega^k + \lambda f))_{i,j}|\}} \tag{8}$$

应用梯度投影算法的基本结果, 文献 [17] 证明了当 $0 < \tau \leq 1/4$ 时, 式 (8) 是收敛的. 试验表明当 $\tau = 0.248$ 时会有更好的数值表现. 不管是半隐式梯度下降算法 (7) 还是梯度投影算法 (8) 都需要函数值在每次迭代时单调下降. 但当目标函数病态时, 收敛到最优值的速度常常会变慢. 这可以解释为用最速下降法求解无约束优化问题, 初始的几步收敛得比较快, 这是为什么 Chambolle 梯度投影算法能快速地收敛于中等精度的原因. 但当接近于最优解时, 算法会出现“锯齿”现象, 导致收敛速度变慢. 采用 BB 步长可以有效地解决上述问题. 特别是对于二次函数, 在问题病态时, 交替使用 BB 步长会克服问题的病态性, 更快地收敛于最优解, 详细分析见文献 [18]. 许多文献表明, 在高精度要求时采用自适应 BB 步长会有更好的数值表现, 本文数值试验部分也验证了这一点.

本节提出一个求解问题 (4) 的非单调自适应步长梯度投影算法, 它有效地改进上述算法中式 (7) 和 (8) 收敛慢的缺点. 问题 (4) 可以看作是在闭凸集上求解一个可微凸函数的最小化问题. 首先考虑无约

束优化问题 $\min_{\omega \in \mathbb{R}^n} F(\omega)$, 梯度方法的迭代过程为 $\omega^{k+1} = \omega^k - \mu^k g^k, k = 0, 1, \dots, g^k = \nabla F(\omega^k)$, 其中步长 $\mu^k > 0$ 通过合适的选择规则确定.

Barzilai 等提出一个独创性的梯度方法^[19], 步长 $\mu_{BB}^k (k > 0)$ 由下面两式决定:

$$\mu_{BB1}^k = \frac{\langle s^{k-1}, s^{k-1} \rangle}{\langle y^{k-1}, s^{k-1} \rangle} \quad (9)$$

$$\mu_{BB2}^k = \frac{\langle s^{k-1}, y^{k-1} \rangle}{\langle y^{k-1}, y^{k-1} \rangle} \quad (10)$$

其中 $s^{k-1} = \omega^k - \omega^{k-1}, y^{k-1} = g^k - g^{k-1}$. 实际上, μ_{BB}^k 是以下两个近似割线方程

$$\mu_{BB1}^k = \arg \min_{\mu \in \mathbb{R}} \|\mu s^{k-1} - y^{k-1}\|$$

和

$$\mu_{BB2}^k = \arg \min_{\mu \in \mathbb{R}} \|s^{k-1} - \mu^{-1} y^{k-1}\|$$

的最优解. 它是投影拟牛顿法的一种快速逼近算法. 由于其形式简单以及数值表现突出, 故在工程应用中得到了极大的关注.

令 $\omega \in Y, F(\omega) := \frac{1}{2} \|\lambda f + \nabla \cdot \omega\|_X^2$. 那么 $g^k := \nabla F(\omega^k) = -\nabla(\nabla \cdot \omega^k + \lambda f)$. 所以有 $y^{k-1} = -\nabla(\nabla \cdot s^{k-1})$ 和

$$\begin{aligned} \langle y^{k-1}, s^{k-1} \rangle_Y &= \langle -\nabla(\nabla \cdot s^{k-1}), s^{k-1} \rangle_Y = \\ &= \langle -(\nabla \cdot s^{k-1}), -(\nabla \cdot s^{k-1}) \rangle_X \end{aligned}$$

因此 μ_{BB}^k 有如下形式:

$$\mu_{BB1}^k = \frac{\langle s^{k-1}, s^{k-1} \rangle}{\langle y^{k-1}, s^{k-1} \rangle} = \frac{\|s^{k-1}\|_Y^2}{\|\nabla \cdot s^{k-1}\|_X^2} \quad (11)$$

$$\mu_{BB2}^k = \frac{\langle s^{k-1}, y^{k-1} \rangle}{\langle y^{k-1}, y^{k-1} \rangle} = \frac{\|\nabla \cdot s^{k-1}\|_X^2}{\|\nabla(\nabla \cdot s^{k-1})\|_Y^2} \quad (12)$$

用式 (11) 和 (12) 代替 Chambolle 算法中的时间步长 τ , 得到两个相应的非单调 Chambolle 投影算法. 即非单调 Chambolle 半隐式梯度投影算法:

$$\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k) := \frac{\omega^k - \mu_{BB}^k g^k}{1 + \mu_{BB}^k |g^k|} \quad (13)$$

和非单调 Chambolle 梯度投影算法:

$$\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k) := \frac{\omega^k - \mu_{BB}^k g^k}{\max\{1, |\omega^k - \mu_{BB}^k g^k|\}} \quad (14)$$

令 $p(g^k)$ 定义如下

$$p(g^k) = \omega^{k+1}(\omega^k, 1, g^k) - \omega^k \quad (15)$$

$\omega^{k+1}(\omega^k, 1, g^k)$ 由式 (13) 和 (14) 定义. 如果 $\|p(g^k)\| = 0$, 那么 ω^k 是一个约束稳定点. 因此, 得到下面的性质.

引理 1. 如果 $\|p(g^k)\| = 0$, 那么 ω^k 满足欧拉-拉格朗日方程 (5).

证明. 当 $\omega^{k+1}(\omega^k, 1, g^k)$ 由式 (13) 计算, 有 $\frac{\omega_{i,j}^k - g_{i,j}^k}{1 + |g_{i,j}^k|} - \omega_{i,j}^k = 0, \forall i, j$. 所以, 欧拉-拉格朗日方程 (5) 成立. 当 $\omega^{k+1}(\omega^k, 1, g^k)$ 由式 (14) 计算, 那么有

$$\omega_{i,j}^k - g_{i,j}^k - \max\{1, |\omega_{i,j}^k - g_{i,j}^k|\} \omega_{i,j}^k = 0, \forall i, j$$

当 $|\omega_{i,j}^k - g_{i,j}^k| \leq 1$ 时, 可以得到 $g_{i,j} = 0$. 否则, $|\omega_{i,j}^k - g_{i,j}^k| > 1$ 且 $g_{i,j} \neq 0$. 因此可以推导出下式

$$\omega_{i,j}^k = \frac{-g_{i,j}^k}{|\omega_{i,j}^k - g_{i,j}^k| - 1}$$

令 $\frac{1}{|\omega_{i,j}^k - g_{i,j}^k| - 1} = t$, 则有 $\omega_{i,j}^k = -tg_{i,j}^k$ 和 $t(t + 1)|g_{i,j}^k| - t = 1$. 又因为 $g_{i,j}^k \neq 0$ 和 $t > 0$, 推出 $t = 1/|g_{i,j}^k|$. 所以, $g_{i,j}^k + |g_{i,j}^k| \omega_{i,j}^k = 0$. 因此等式 (5) 成立. \square

下面给出改进的非单调线搜索梯度投影算法的具体步骤, 记为算法 1.

算法 1. 改进的非单调线搜索梯度投影算法

步骤 1. 给定 $\omega^0 \in Y, M \geq 1, \theta \in (0, 1), \epsilon \in (0, 1),$

$k := 0, 0 < \rho_{\min} < \rho_{\max}$.

步骤 2. 如果 $\|p(g^k)\| = 0$, 停止.

步骤 3. 限制 $\mu_{BB}^k \in [\rho_{\min}, \rho_{\max}]$.

步骤 4. $d^k := \omega^{k+1}(\omega^k, \mu_{BB}^k, g^k) - \omega^k$.

步骤 5. 改进的非单调线搜索:

令 $\gamma \in [0, \frac{-\theta \epsilon \langle g^k, d^k \rangle}{\|d^k\|^2}]$, 如果

$$F(\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k)) \leq \max_{0 \leq j < \min\{k, M-1\}} F(\omega^{k-j}) + \theta \langle g^k, d^k \rangle + \gamma \|d^k\|^2 \quad (16)$$

$\omega^{k+1} := \omega^{k+1}(\omega^k, \mu_{BB}^k, g^k)$, 转步骤 6. 否则 $\sigma \in (0, 1), \mu_{BB}^k = \sigma \mu_{BB}^k$, 转步骤 5.

步骤 6. 令 $k = k + 1$, 转步骤 2.

步骤 3 中的自适应 BB 步长选择算法:

步骤 3.1. 若 $k = 0$, 令 $\mu_0 \in [\rho_{\min}, \rho_{\max}], \tau_1 \in (0, 1), M_\mu$ 为一正整数.

步骤 3.2. 若 $s_{k-1}^T y^{k-1} \leq 0$, 令 $\mu_{BB}^k = \rho_{\max}$, 转步骤 3.6, 否则转步骤 3.3.

步骤 3.3. 求

$$\mu_{BB1}^k = \max\{\rho_{\min}, \min\{\frac{\|s^{k-1}\|_Y^2}{\|\nabla \cdot s^{k-1}\|_X^2}, \rho_{\max}\}\}$$

$$\mu_{BB2}^k = \max\{\rho_{\min}, \min\{\frac{\|\nabla \cdot s^{k-1}\|_X^2}{\|\nabla(\nabla \cdot s^{k-1})\|_Y^2}, \rho_{\max}\}\}$$

步骤 3.4. 若 $\frac{\mu_{BB2}^k}{\mu_{BB1}^k} \leq \tau_k$, 求 $\mu_{BB}^k = \min\{(\mu_{BB2}^k)_j, j = \max\{1, k - M_\mu\}, \dots, k\}, \tau_{k+1} = 0.4\tau_k$, 否则转步骤 3.5.

步骤 3.5. $\mu_{BB}^k = \mu_{BB1}^k, \tau_{k+1} = 1.5\tau_k$, 转步骤 3.4.

步骤 3.6. 求得 μ_{BB}^k , 结束.

定义 $g_s(\omega) = \omega(\omega, \mu_{BB}, g) - \omega$, 其中 $s \in [\rho_{\min}, \rho_{\max}]$. 由文献 [20] 中的引理 2.1 可知,

$$\langle g(\omega), g_s(\omega) \rangle \leq -\frac{1}{s} \|g_s(\omega)\|^2 \leq -\frac{1}{\rho_{\max}} \|g_s(\omega)\|^2 \quad (17)$$

如果 ω^k 不是稳定点, 那么 $\langle g^k, d^k \rangle < -\|d^k\|^2/\rho_{\max}$, 因此搜索方向是下降方向. 如果 $\gamma = 0$, 式 (16) 就是标准的非单调线搜索. 再由文献 [20] 中的定理 2.3 知, 只要 $\mu_{BB}^k \in [\rho_{\min}, \rho_{\max}]$, 算法 1 是收敛的. 下面给出一般情形下算法 1 的收敛性.

定理 2. 算法 1 是适定的, 且序列 $\{\omega^k\}$ 的任意聚点 ω^* 是式 (4) 的约束稳定点.

证明. 由文献 [20] 中的引理 2.2 (ii) 知, 对任意 $\alpha_{BB}^k \in [\rho_{\min}, \rho_{\max}]$ 有

$$F(\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k)) - \max_{0 \leq j \leq \min\{k, M-1\}} F(\omega^{k-j}) \leq F(\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k)) - F(\omega^k) \leq \theta \langle g^k, d^k \rangle$$

故

$$F(\omega^{k+1}(\omega^k, \mu_{BB}^k, g^k)) - \max_{0 \leq j \leq \min\{k, M-1\}} F(\omega^{k-j}) \leq \theta \langle g^k, d^k \rangle + \gamma \|d^k\|^2$$

再由 γ 的取值可得

$$\theta \langle g^k, d^k \rangle + \gamma \|d^k\|^2 < 0$$

所以式 (16) 是非单调下降, 步长可以有限步得到. 因此, 算法 1 是适定的.

令 ω^* 是 $\{\omega^k\}$ 的聚点. 考虑下面两种情况:

情况 1. 如果 $\inf \alpha_{BB}^k = 0$, 那么存在一个子序列 $\{\omega^k\}_K$ 使得

$$\lim_{k \in K} \alpha_{BB}^k = 0$$

在这种情况下, 存在充分大的 \bar{k} , 使得当 $k \geq \bar{k}, k \in K$, 存在 $\sigma_k \in (0, 1)$, $\psi^k = \alpha_{BB}^k/\sigma_k$ 不满足式 (16), 即

$$F(\omega^{k+1}(\omega^k, \psi^k, g^k)) > \max_{0 \leq j \leq \min\{k, M-1\}} F(\omega^{k-j}) + \theta \langle g^k, d^k \rangle + \gamma \|d^k\|^2 \geq F(\omega^k) + \theta \langle g^k, d^k \rangle$$

剩下的证明类似于文献 [20] 中定理 2.3, 此处不再详细给出, 有兴趣的读者可以参考文献 [20]. 因此, ω^* 是一个约束稳定点.

情况 2. 假设 $\inf \alpha_{BB}^k \geq \beta > 0$. 反证法. ω^* 不是一个约束稳定点, 那么对所有的 $k > 0$

$\langle g^k, d^k \rangle \leq -\zeta, \zeta > 0$. 由文献 [21] (p.709) 知, 可以得到一个单调非增的序列 $\{f(\omega^{l(k)})\}$. 其中 $l(k)$ 是使得 $k - \min\{k, M-1\} \leq l(k) \leq k$ 和

$$F(\omega^{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} f(\omega^{k-j})$$

成立的整数. 再由式 (16) 可得, 对于 $k > M-1$ (见文献 [21]),

$$F(\omega^{l(k)}) \leq F(\omega^{l(l(k)-1)}) + \theta \langle g(x^{l(k)-1}), d^{l(k)-1} \rangle + \gamma \|d^{l(k)-1}\|^2$$

又因为 $\gamma \in [0, \frac{\epsilon \theta g^k d^k}{\|d^k\|^2}]$, 故可推出

$$\begin{aligned} F(\omega^{l(k)}) - F(\omega^{l(l(k)-1)}) &\leq \theta \langle g(x^{l(k)-1}), d^{l(k)-1} \rangle + \gamma \|d^{l(k)-1}\|^2 \leq \\ &\theta \langle g(x^{l(k)-1}), d^{l(k)-1} \rangle - \\ &\frac{\epsilon \theta g(x^{l(k)-1})^T d^{l(k)-1}}{\|d^{l(k)-1}\|^2} \|d^{l(k)-1}\|^2 \leq \\ &\theta(1-\epsilon) \langle g(x^{l(k)-1}), d^{l(k)-1} \rangle \leq \\ &-\theta(1-\epsilon)\zeta \end{aligned}$$

当 $k \rightarrow \infty$ 时, $F(\omega^{l(k)}) \rightarrow -\infty$, 与 $F(\omega)$ 有界矛盾. 因此, ω^* 是一个约束稳定点. \square

定理 3 是文献 [22] 中的命题 1, 该性质说明由对偶问题的解可以得到原问题的解.

定理 3. 令 $\{\omega^k\}$ 是任意的序列, $\omega^k \in \mathcal{K}, k = 1, 2, \dots, \mathcal{K} = \{\omega : |\omega| \leq 1\}$ 且 $\{\omega^k\}$ 的所有聚点是式 (4) 的稳定点. 那么当 $k \rightarrow \infty$ 时, 序列 $u^k := f + \frac{1}{\lambda} \nabla \cdot \omega^k$ 收敛于式 (1) 的唯一解 \bar{u} .

2 数值结果

本节给出算法 1 的数值表现. 仿真实验在内存为 2GB 处理器为 i3 的个人电脑上进行. 在两个实验中, Matlab 版本为 R2011a, 高斯噪音由 Matlab 中的加噪函数 imnoise 产生, 方差是 0.01. 采用均方根误差 (Root mean square error, RMSE) 评价算法恢复的效果, 定义如下

$$RMSE = \frac{\|u - u_k\|_2}{\|u\|_2}$$

其中 u 是真实的图像, u_k 是算法得到的图像.

2.1 和 Chambolle 梯度投影算法比较

本节将算法 1 和文献 [7-8] 中的方法进行对比. 在算法 1 中, 参数设置如下, $\alpha_0 = 1, M = 5, \theta = 10^{-5}, \rho_{\max} = 10^{10}, \rho_{\min} = 10^{-10}, \sigma = 0.5$. 保真项 $\lambda = 0.053$.

表 1 数值结果
Table 1 Numerical results

Image	Algorithm	Niter	Time (s)	RMSE
Shape (128 × 128)	Chambolle	737	2.06	0.0739
	Chambolle1	592	1.86	0.0738
	MChambolle	196	0.81	0.0412
	GPSSABB	182	0.70	0.0414
	MGPSSABB	130	0.61	0.0414
Lena (256 × 256)	Chambolle	352	5.79	0.0829
	Chambolle1	272	4.73	0.0827
	MChambolle	129	3.62	0.0221
	GPSSABB	152	4.13	0.0214
	MGPSSABB	85	2.34	0.0214
Peppers (512 × 512)	Chambolle	389	35.80	0.0525
	Chambolle1	298	26.27	0.0520
	MChambolle	124	19.77	0.0520
	GPSSABB	124	19.67	0.0079
	MGPSSABB	92	15.79	0.0079
Boat (1024 × 1024)	Chambolle	388	141.80	0.0078
	Chambolle1	300	117.05	0.0078
	MChambolle	140	101.66	0.0078
	GPSSABB	156	107.90	0.0078
	MGPSSABB	97	67.16	0.0078

测试如下算法:

- 1) 文献 [7] 算法: 半隐式梯度下降算法, $\tau = 0.248$, 记为 Chambolle.
- 2) 文献 [8] 算法: 梯度投影下降算法, $\tau = 0.248$, 记为 Chambolle1.
- 3) MChambolle: 算法 1 中自适应步长选择算法, $\omega^{k+1}(\cdot)$ 由式 (13) 计算, $\gamma = 0.5$.
- 4) GPSSABB: 算法 1 中自适应步长选择算法, $\omega^{k+1}(\cdot)$ 由式 (14) 计算, $\gamma = 0$.
- 5) MGPSSABB: 算法 1 中自适应步长选择算法, $\omega^{k+1}(\cdot)$ 由式 (14) 计算, $\gamma = 0.5$. 五种方法的停止准则为

$$\|p(g^k)\| \leq 10^{-6} \|p(g^0)\|$$

为了公平测试算法的速率, 每幅图像进行 10 次试验, 表 1 给出迭代步数和 CPU 时间 (秒) 的 10 次平均数值结果. 可以看出, 四种方法恢复的 RMSE 值相差不大, 但算法 1 有效地改善了 Chambolle 算法的收敛速度. 图 1 给出了 Boat (1024 × 1024) 恢复的结果. 图 2 说明非单调下降算法有效地改善原始算法的数值表现, 特别是对高精度要求和大尺寸的图像效果更好.



图 1 Boat (1024 × 1024) 去噪结果
Fig. 1 Denoising results of Boat (1024 × 1024)

2.2 和梯度投影类算法比较

文献 [22] 给出求解式 (4) 的一系列 BB 梯度投影算法且在基于对偶间隙的基础上给出算法的一个停止准则. 下面给出文献 [22] 中基于对偶间隙的停止准则.

由 TV 范数的定义, ROF 模型有如下形式:

$$\min_u \max_{|\omega| \leq 1} \int_{\Omega} -u \nabla \cdot \omega + \frac{\lambda}{2} \|u - f\|^2$$

利用文献 [23] 中命题 2.4 的 min-max 定理, 问题 (1) 可转化为:

$$\max_{|\omega| \leq 1} \min_u \int_{\Omega} u \nabla \cdot \omega + \frac{\lambda}{2} \|u - f\|^2 \quad (18)$$

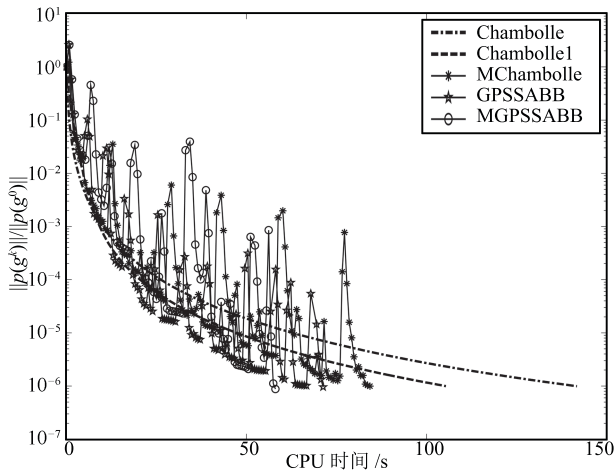


图 2 Boat (1024 × 1024) 梯度投影的相对范数 vs. CPU 时间 (秒)

Fig. 2 Relative norm of projected gradient vs. CPU time (s) of Boat (1024 × 1024)

式 (18) 中内层最小化问题的精确解为:

$$u = f + \frac{1}{\lambda} \nabla \cdot \omega \quad (19)$$

将式 (19) 代入式 (18) 中, 可得对偶问题:

$$\max_{|\omega| \leq 1} D(\omega) := \frac{\lambda}{2} \left[\|f\|^2 - \left\| f + \frac{1}{\lambda} \nabla \cdot \omega \right\|^2 \right] \quad (20)$$

对偶间隙定义如下:

$$G(u, \omega) = P(u) - D(\omega) = \int_{\Omega} |\nabla u| + \frac{\lambda}{2} \|u - f\|^2 - \frac{\lambda}{2} \left[\|f\|^2 - \left\| f + \frac{1}{\lambda} \nabla \cdot \omega \right\|^2 \right] = \int_{\Omega} (|\nabla u| - \nabla u \cdot \omega) + \frac{\lambda}{2} \left\| \frac{1}{\lambda} \nabla \cdot \omega + f - u \right\|^2$$

如果 u 和 ω 是可行解, 那么

$$0 \leq P(u) - O \leq G(u, \omega)$$

$$0 \leq O - D(\omega) \leq G(u, \omega)$$

其中 O 为原对偶最优值. 如果 (u, ω) 是原对偶问题的最优解, 那么

$$G(u, \omega) = \int_{\Omega} (|\nabla u| - \nabla u \cdot \omega)$$

下面具体分析当 $\omega = \omega^k$ 满足如下停止准则时

$$\frac{G(u, \omega)}{|P(u)| + |D(\omega)|} \leq Tol \quad (21)$$

由算法 1 得到的 u 逼近最优解 \bar{u} , 其中 Tol 是很小的正整数. 从式 (3), 可以推出

$$u - \bar{u} = \left(f + \frac{1}{\lambda} \nabla \cdot \omega \right) - \left(f + \frac{1}{\lambda} \nabla \cdot \bar{\omega} \right) = \frac{1}{\lambda} (\nabla \cdot \omega - \nabla \cdot \bar{\omega})$$

又因为 $|\nabla \bar{u}| = \nabla \bar{u} \cdot \bar{\omega}$ 和 $|\nabla u| \geq \nabla u \cdot \omega$ ($|\omega| \leq 1$) 故

$$\begin{aligned} \lambda \|u - \bar{u}\|_2^2 &= \int_{\Omega} (u - \bar{u})(\nabla \cdot \omega - \nabla \cdot \bar{\omega}) = \\ &= \int_{\Omega} u \nabla \cdot \omega - \int_{\Omega} u \nabla \cdot \bar{\omega} - \int_{\Omega} \bar{u} \nabla \cdot \omega + \int_{\Omega} \bar{u} \nabla \cdot \bar{\omega} = \\ &= \int_{\Omega} -\nabla u \cdot \omega + \int_{\Omega} \nabla u \cdot \bar{\omega} + \int_{\Omega} \nabla \bar{u} \cdot \omega - \int_{\Omega} |\nabla \bar{u}| \leq \\ &= \int_{\Omega} -\nabla u \cdot \omega + \int_{\Omega} |\nabla u| + \int_{\Omega} |\nabla \bar{u}| (|\omega| - 1) \leq \\ &= \int_{\Omega} (|\nabla u| - \nabla u \cdot \omega) = G(u, \omega) \end{aligned}$$

所以

$$\|u - \bar{u}\| \leq \sqrt{\frac{G(u, \omega)}{\lambda}} \leq \sqrt{\frac{Tol(|P(u)| + |D(\omega)|)}{\lambda}}$$

本节给出算法在停止准则 (21) 下的数值表现, 并与文献 [22] 中一些梯度投影算法比较. 测试图像为 Cameraman (256 × 256), Barbara (512 × 512). 在试验中, 保真参数 $\lambda = 0.045$, $\rho_{\max} = 10^5$, $\rho_{\min} = 10^{-5}$, $M = 5$, $\theta = 10^{-4}$, $\alpha_0 = 1$, $\sigma = 0.5$, $M_{\mu} = 2$. 测试如下算法.

1) 文献 [7] 算法: 半隐式梯度下降算法, $\tau = 0.248$, 记为 Chambolle.

2) 文献 [22] 的一些梯度投影算法: GPCL, GPLS, GPBBM, GPBBM(3), SQPBBM, GPBB-safe, GPBBNM, GPABB.

3) GPSSABB: 算法 1 中自适应步长选择算法, $\omega^{k+1}(\cdot)$ 由式 (14) 计算, $\gamma = 0$.

4) MGPSSABB: 算法 1 中自适应步长选择算法, $\omega^{k+1}(\cdot)$ 由式 (14) 计算, $\gamma = 0.5$.

图 3 分别给出 $Tol = 10^{-4}$ 时 Cameraman 和 $Tol = 10^{-6}$ 时 Barbara 图像恢复的结果. 可以看出, 算法 1 能够有效地去除噪声得到清晰的图像. 表 2 和 3 给出 $Tol = 10^{\{-2, -3, -4, -6\}}$ 的数值结果. 其中 GPBBNM 表示在标准非单调线搜索下仅使用 μ_{BB1}^k 的梯度投影算法. GPABB 是另一种形式的交替使用 BB 步长梯度投影算法. GPBBNM 和 GPABB 算法是文献 [22] 中表现最好的算法, 具体形式可以参考文献 [22]. 从这两个表中可以发现, $Tol \leq 10^{-4}$ 时, 新算法优势并不明显, 甚至稍微慢于其他算法, 如 GPABB 和 GPBBNM. 但 $Tol = 10^{-6}$ 时,

表 2 Cameraman (256 × 256) 迭代步数和 CPU 时间 (秒)
Table 2 Number of iterations (Iter) and CPU time (s) of Cameraman (256 × 256)

Algorithms	$Tol = 10^{-2}$		$Tol = 10^{-3}$		$Tol = 10^{-4}$		$Tol = 10^{-6}$	
	Iter	Time (s)	Iter	Time (s)	Iter	Time (s)	Iter	Time (s)
Chambolle	27	0.45	163	3.18	822	16.04	14 625	271.89
GPCL	32	0.59	114	1.98	549	10.19	10 116	187.28
GPLS	18	0.53	132	5.38	607	26.22	12 100	550.32
GPBBM	20	0.56	128	3.24	596	16.99	11 032	295.31
GPBBM(3)	20	0.58	72	1.78	292	7.53	3 186	79.15
SQPBBM	14	0.50	43	1.95	167	7.66	2 532	105.18
GPBBsafe	16	0.33	47	1.05	165	4.23	2 398	73.68
GPBBNM	16	0.34	48	1.06	162	3.42	2 974	62.19
GPABB	16	0.45	47	1.23	179	4.68	2 276	49.45
GPSSABB	13	0.41	48	1.36	146	4.31	1 372	38.50
MGPSSABB	13	0.37	47	1.33	129	3.67	678	18.90

表 3 Barbara (512 × 512) 迭代步数和 CPU 时间 (秒)
Table 3 Number of iterations (Iter) and CPU time (s) of Barbara (512 × 512)

Algorithms	$Tol = 10^{-2}$		$Tol = 10^{-3}$		$Tol = 10^{-4}$		$Tol = 10^{-6}$	
	Iter	Time (s)	Iter	Time (s)	Iter	Time (s)	Iter	Time (s)
Chambolle	27	3.04	131	14.27	536	60.47	8 583	939.73
GPCL	24	2.78	79	9.31	331	37.10	5 854	641.63
GPLS	36	6.61	90	20.90	319	77.84	5 532	1 316.20
GPBBM	20	3.03	84	13.32	340	52.10	6 096	872.62
GPBBM(3)	20	2.89	74	11.31	229	34.94	2 588	365.88
SQPBBM	14	2.75	33	7.71	106	23.20	1 530	325.25
GPBBsafe	15	2.09	40	5.76	118	18.75	1 494	290.38
GPBBNM	15	2.42	41	5.55	116	15.69	1 467	194.15
GPABB	14	2.43	35	5.91	125	21.23	1 123	180.51
GPSSABB	12	2.03	34	5.97	104	19.25	878	159.95
MGPSSABB	14	2.39	37	6.26	90	15.55	431	86.31

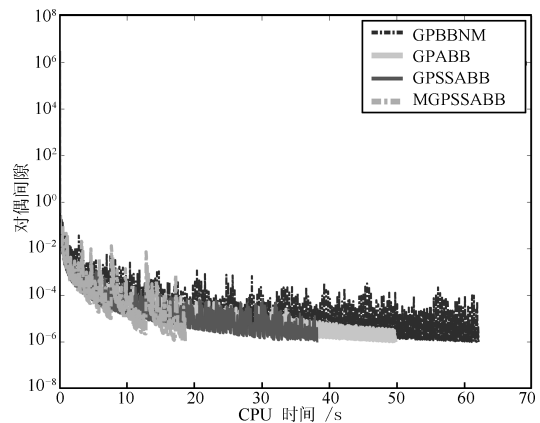


图 3 Cameraman ($Tol = 10^{-4}$) 和 Barbara ($Tol = 10^{-6}$) MGPSSABB 的去噪结果

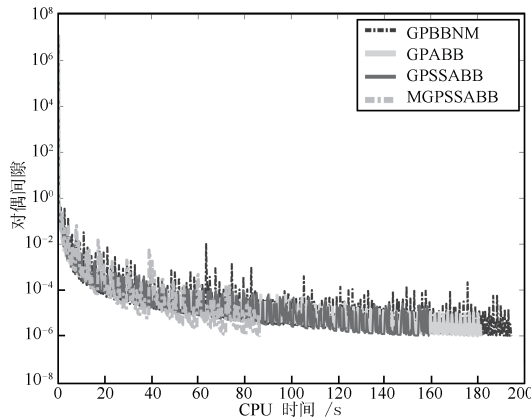
Fig. 3 Denoising results of MGPSSABB of Cameraman ($Tol = 10^{-4}$) and Barbara ($Tol = 10^{-6}$)

MGPSSABB 的优势明显地表现出来, 迭代步数和 CPU 时间比 GPABB 和 GPBBNM 大大减少. 图 4 给出了对偶间隙随时间变化的曲线. 在算法满足停止准则终止时, 从图 4 可以看出新算法 CPU 时间远少

于 GPABB 和 GPBBNM 算法的 CPU 时间. 这也证实了算法 1 在采用自适应 BB 步长和改进的线搜索下, 能够更快地收敛于最优解, 特别是精确恢复大尺寸图像和高精度要求时.



(a) Cameraman
(a) Cameraman



(b) Barbara

(b) Barbara

图 4 四个算法的对偶间隙 vs. CPU 时间 (秒) ($Tol = 10^{-6}$)Fig. 4 Duality gap vs. CPU time(s) for four algorithms ($Tol = 10^{-6}$)

通过图 2 和图 4 发现, 相对范数和对偶间隙在迭代过程中不是单调下降. 而这个非单调性是文献 [16] 指出的一个重要特征, 它可能使算法更快地收敛于最优解.

最后, 这两节的数值试验表明新提出的自适应步长梯度投影算法是有效的, 优于其他几种梯度投影算法.

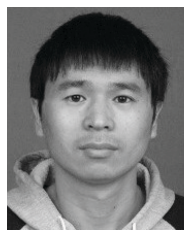
3 结论

针对全变差图像去噪问题, 本文提出了一个自适应 BB 步长梯度投影算法, 用著名的 BB 步长代替 Chambolle 梯度投影算法中的时间常数步长, 并在改进的非单调线搜索下证明算法的全局收敛性. 数值结果表明, 新提的算法有效地改善了原始方法. 从表 1 中数值结果可以看出, 特别是对于大尺寸图像, 新算法运行的 CPU 时间大约是 Chambolle 半隐式梯度投影方法的 CPU 时间的一半. 并与其他梯度投影方法做比较, 其数值表现也有明显优势.

References

- Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 1992, **60**(1-4): 259-268
- Aubert G, Kornprobst P. *Mathematical Problems in Image Processing*. Berlin: Springer, 2002.
- Chan T, Shen J H. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Philadelphia: SIAM, 2005.
- Wang Shi-Yan, Yu Hui-Min. Motion segmentation model based on total variation and split Bregman algorithm. *Acta Automatica Sinica*, 2015, **41**(2): 396-404
(王诗言, 于慧敏. 基于全变分的运动分割模型及分裂 Bregman 算法. 自动化学报, 2015, **41**(2): 396-404)
- Zhang Rui, Feng Xiang-Chu, Wang Si-Qi, Chang Li-Hong. A sparse gradients field based image denoising algorithm via non-local means. *Acta Automatica Sinica*, 2015, **41**(9): 1542-1552
(张瑞, 冯象初, 王斯琪, 常莉红. 基于稀疏梯度场的非局部图像去噪算法. 自动化学报, 2015, **41**(9): 1542-1552)
- Chan T F, Golub G H, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 1999, **20**(6): 1964-1977
- Chambolle A. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 2004, **20**(1): 89-97
- Chambolle A. Total variation minimization and a class of binary MRF models. In: *Proceedings of the 5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. St. Augustine, FL, USA: Springer, 2005. 136-152
- Ren Fu-Qian, Qiu Tian-Shuang. Blurred image restoration method based on second-order total generalized variation regularization. *Acta Automatica Sinica*, 2015, **41**(6): 1166-1172
(任福全, 邱天爽. 基于二阶广义全变差正则项的模糊图像恢复算法. 自动化学报, 2015, **41**(6): 1166-1172)
- Zhang B X, Zhu Z B, Li S A. A modified spectral conjugate gradient projection algorithm for total variation image restoration. *Applied Mathematics Letters*, 2014, **27**: 26-35
- Xiao Y H, Wu S Y, Qi L Q. Nonmonotone Barzilai-Borwein gradient algorithm for l_1 regularized nonsmooth minimization in compressive sensing. *Journal of Scientific Computing*, 2014, **61**(1): 17-41
- Loris I, Bertero M, De Mol C, Zanella R, Zanni L. Accelerating gradient projection methods for l_1 -constrained signal recovery by steplength selection rules. *Applied and Computational Harmonic Analysis*, 2009, **27**(2): 247-254
- Huang Y K, Liu H W, Zhou S. An efficient monotone projected Barzilai-Borwein method for nonnegative matrix factorization. *Applied Mathematics Letters*, 2015, **45**: 12-17
- Xiao Y H, Jin Z F. An alternating direction method for linear-constrained matrix nuclear norm minimization. *Numerical Linear Algebra with Applications*, 2012, **19**(3): 541-554
- Yu G H, Qi L Q, Dai Y H. On nonmonotone Chambolle gradient projection algorithms for total variation image restoration. *Journal of Mathematical Imaging and Vision*, 2009, **35**(2): 143-154
- Dai Y H, Fletcher R. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. *Numerische Mathematik*, 2005, **100**(1): 21-47

- 17 Bertsekas D P. *Nonlinear Programming*. Nashua: Athena Scientific, 1999.
- 18 Frassoldati G, Zanghirati G, Zanni L. New adaptive step-size selections in gradient methods. *Journal of Industrial and Management Optimization*, 2008, **4**(2): 299–312
- 19 Barzilai J, Borwein J M. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 1988, **8**(1): 141–148
- 20 Birgin E G, Martínez J M, Raydan M. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 2000, **10**(4): 1196–1211
- 21 Grippo L, Lampariello F, Lici S. A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, 1986, **23**(4): 707–716
- 22 Zhu M Q, Wright S J, Chan T F. Duality-based algorithms for total-variation image restoration. *Computational Optimization and Applications*, 2010, **47**(3): 377–400
- 23 Ekeland I, Témam R. *Convex Analysis and Variational Problems*. Philadelphia: SIAM, 1999.



张本鑫 桂林电子科技大学电子工程与自动化学院博士研究生. 主要研究方向为最优化方法和变分法在图像处理中的应用. E-mail: zbx913@163.com
(**ZHANG Ben-Xin** Ph.D. candidate at the School of Electronic Engineering and Automation, Guilin University of Electronic Technology. His research interest covers optimization algorithm, variational method and their applications in image processing.)



朱志斌 桂林电子科技大学数学与计算科学学院教授. 2004 年获西安交通大学理学博士学位. 主要研究方向为最优化方法及其在图像处理和反问题中的应用. 本文通信作者.
E-mail: optimization_zhu@163.com
(**ZHU Zhi-Bin** Professor at the School of Mathematics and Computing Science, Guilin University of Electronic Technology. He received his Ph. D. degree in applied mathematics from Xi'an Jiaotong University in 2004. His research interest covers optimization and their applications in image processing and its inverse problem. Corresponding author of this paper.)